DOCUMENT RESUME

ED 318 777 TM 014 877

AUTHOR Sudweeks, Richard R.; And Others
TITLE Test-Wise Responses of Third-, Fifth-, and Sixth-Grade Students to Clued and Unclued Multiple-Choice Science Items.
PUB DATE 89
NOTE 37p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Age Differences; Child Development; Cognitive Processes; Comparative Testing; *Cues; Elementary Education; *Elementary School Students; Grade 3; Grade 5; Grade 6; Individual Development; Intermediate Grades; *Multiple Choice Tests; Preadolescents; Problem Solving; Science Education; *Science Tests; Testing Problems; *Test Wiseness
IDENTIFIERS Embedded Clues

ABSTRACT

Testwise responses of grade-school students to multiple-choice science items and the intercorrelations between different testwiseness skills were examined in this document. A total of 134 third-grade, 133 fifth-grade, and 143 sixth-grade students from two elementary schools in a large suburban school district in Utah responded to 60 science items on tests developed specifically for each of these grades. Thirty of the items contained embedded clues intended to elicit one of five separate testwise behaviors. The other 30 science items were identical to the original items, but did not contain the embedded clues. Half of the students completed the items without embedded clues before the items with clues, while the rest of the students completed the items with clues first. The presence of embedded clues made no difference in the performance of third graders, but both fifth graders and sixth graders demonstrated an ability to use the embedded clues to improve their performance. Testwiseness skills related to verbal associations between the stem and correct response and to absurd response options were observed in fifth and sixth graders. The next skill to appear by the fifth grade was a skill related to recognizing synonymous or overlapping response options. None of the skills were intercorrelated among the fifth graders, but among sixth graders, the verbal associations skill and absurd options skill were moderately correlated. Testwiseness skills appear to develop independently of each other as students develop more complex cognitive processes. Testwiseness skills do not develop in all students at the same time. Sample test items are included, and 12 tables present study data. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED318727

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
GARY A. PETERSEN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Test-Wise Responses of Third-, Fifth-, and Sixth-Grade Students to Clued and Unclued Multiple-Choice Science Items

Richard R Sudweeks, J. Hugh Baird, and Gary A. Petersen
Brigham Young University

Running Head: TEST-WISE RESPONSES

BEST COPY AVAILABLE

TM014877

Abstract

This study investigated the test-wise responses of grade school students to multiple-choice science items and examined the intercorrelations between different test-wiseness skills. A total of 134 third-, 133 fifth-, and 143 sixth-grade students responded to 60 science items on tests developed specifically for each of these grades. Thirty of the items contained embedded clues intended to elicit one of five separate test-wise behaviors. The other 30 science items were identical to the original items except they did not contain the embedded clues. Fifth and sixth graders also responded to an additional set of 30 pseudo-items made up of imaginary or fictitious science content.

Results of the study showed that presence of embedded clues made no difference in the performance of third graders, but both fifth and sixth graders demonstrated an ability to use the embedded clues to improve their performance. Test-wiseness skills related to (a) verbal associations between the stem and correct response and to (b) absurd response options were observed in fifth- and sixth-grade students. The next skill to appear by the fifth grade was a skill related to recognizing synonymous or overlapping response options. None of the test-wiseness skills were intercorrelated among fifth-grade students, but among sixth-grade students the verbal associations skill and the absurd options skill were moderately correlated.

The findings suggest that test-wiseness skills develop independently of each other as students develop more complex cognitive processes. Moreover, test-wiseness skills do not develop in all students at the same time.

Test-Wise Responses of Third-, Fifth-, and Sixth-Grade Students
to Clued and Unclued Multiple-Choice Science Items

Test-wiseness is defined as an examinee's "capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score." (Millman, Bishop, & Ebel, 1965, p. 707.) This capacity is not limited to any specific type of content; rather, it is assumed to be generalizable across different subject-matter areas. For example, recent studies have examined test-wiseness in such subjects as: reading (Parrish, 1982; Scruggs, Bennion, and Lifson, 1985); English, history, mathematics, and science (Slakter, Koehler, and Hampton, 1970); and American History (McMorris, Brown, Snyder, and Pruzek, 1972).

A major focus of test-wiseness research since 1979 has been to investigate issues related to how students can be trained to develop effective test-taking skills. In other words, the focus of this research has been on test-taking skills rather than on test-wiseness skills. Test-taking skills refer to strategies such as making appropriate use of time during a test and paying attention to instructions.

On the other hand, test-wiseness skills refer to strategies that are useful in the context of specific test items and relate to content or structural flaws in the items. Much of the earlier test-wiseness research focused on six of these strategies. The six strategies include:

1. Absurd Options: recognizing response options that are inconsistent with the item stem.

2. Synonymous or Overlapping Options: identifying options that imply the correctness of one another.

3. Verbal Associations: recognizing options that resemble some aspect of the item stem.

4. Longer or More Detailed Options: identifying options that are longer or contain more detailed information than other options.

5. Grammatical Inconsistencies: identifying options that are not grammatically consistent with the stem.

6. Specific Determiners: recognizing options that contain specific determiners, words such as always, never, and only.

Problem

Although much research has been completed on test-wiseness, additional research is needed to investigate the prevalence and development of test-wiseness among elementary school students. The recent focus on training students to be test-wise has been directed primarily at students in junior high or high school and college. Very little research examining test-wiseness among grade school children has been completed. Instead, research among these younger students has focused on issues related to teaching test-taking skills and the effects of such training. Scruggs, White, and Bennion (1986, p. 70) define test-taking skills as "skills that enable students to demonstrate more fully the knowledge that they have (e.g., being familiar with the question/answer format, knowing the meaning of vocabulary used in directions, being able to pace oneself appropriately during timed tests)." They analyzed 24 studies that had "investigated the effect of teaching test-taking skills for standardized, group-administered achievement tests" to grade school children. In each study, it was test-taking skills rather than test-wiseness skills among elementary school students that were investigated.

However, in one study that did investigate test-wiseness among older, grade school students, Slakter, Koehler, and Hampton (1970) reported evidence that fifth-grade students demonstrated an ability to eliminate options on a multiple-choice test which were known to be incorrect and to then choose from the smaller set of remaining options (absurd options) and also the ability to select options that resembled the item stem in some way (verbal associations), behaviors associated with test-wiseness. They reported that other test-wise behaviors were not apparent in the examinees they studied until the eighth grade. With the exception of this one study, little research was found that indicated what specific test-wiseness skills students in elementary schools have and have not acquired and in what order the different skills develop.

School districts around the country continue to use standardized tests to assess the achievement of their students. State legislatures and state departments of education are also implementing testing programs in their states. This testing is not limited to students in high school but is becoming more and more prevalent among students in elementary schools.

Professional test developers generally attempt to construct tests that are not susceptible to test-wiseness. How-to-do-it text books and other articles on test construction (e.g., Ebel & Frisbie, 1986; Gronlund, 1985; Haladyna & Downing, 1989; Mehrens & Lehmann, 1984; and Popham, 1981) generally include prescriptive item-writing guidelines that are intended to reduce the likelihood of examinees obtaining high scores simply by using sophisticated test-taking strategies. The guidelines are written to apply not only to tests for junior high, high school, or college students but for tests designed for elementary school students as well.

There is evidence to indicate that test-wiseness skills are prevalent among students in high school and college. However, there is little empirical evidence indicating the extent to which test-wise behaviors are present among younger students. Furthermore, if test-wiseness skills are present among students in elementary schools, it is not known which skills are present. The literature on the cognitive development of children suggests that children develop more abstract cognitive processes as they move from childhood to adolescence (Brainerd, 1978; Case, 1985; and Siegler, 1986). Consequently, test-wiseness skills might be expected to develop as children move toward adolescence.

The purpose of this study was to determine to what extent elementary school children have developed test-wise behaviors and what specific types of test-wiseness, if any, they have acquired. The study was limited to investigating five test-wiseness skills: (a) absurd options, (b) synonymous or overlapping options, (c) verbal associations, (d) longer or more detailed options, and (e) grammatical inconsistencies.

Research Questions

The following research questions were investigated:

1. How is the development of the various test-wiseness skills related to students' grade-level and sex? Which skills appear to develop earliest, and which ones appear to develop later?

2. Which test-wiseness skills are most highly inter-correlated?

3. To what extent are the various test-wiseness skills manifest in clued forms of a science test correlated with performance on unclued versions of the same scales?

Methods

Subjects

A convenience sample of 134 third-grade students (67 boys, 60 girls, and 7 unclassified), 133 fifth-grade students (67 boys, 65 girls, and 1 unclassified), and 143 sixth-grade students (88 boys and 55 girls) from two elementary schools in a large suburban school district in Utah participated in this experiment. Due to limited resources, it was possible to test students in only three grades. Third graders were included instead of fourth graders because third graders are younger and not as likely to be entering adolescence.

Instrumentation

Separate science tests were developed for third-, fifth-, and sixth-grade students. For each grade, the test included 30 items selected from the end-of-level science tests prepared for second-, fourth-, or fifth-grade students (Utah State Board of Education, 1987). Items were not selected because of their science content but because of the ease with which clues could be added. The science content for each grade-level covered a wide range of science topics specified in the state core curriculum.

Test developers had previously carefully reviewed each item in the end-of-level tests for each grade in an attempt to ensure that no overt clues were included that might elicit test-wise behaviors. An alternate version of the test was prepared for each grade which included the same 30 items. However, a clue designed to elicit one of the five test-wise behaviors was embedded in each item in either the stem or one or more of the response options.

In addition, a third section was included in the tests prepared for the fifth and sixth graders. This third section consisted of 30 pseudo-items. Pseudo-items were items

having fictitious content and a contrived correct answer based on one of the five test-wiseness skills being measured. There was no way to logically identify the correct answer without using the specific test-wiseness skill that was being measured. However, five of the pseudo-items contained factual content unrelated to science and not generally taught to students in fifth or sixth grade. Consequently, to logically identify the correct answer to any of these five questions, students had to rely on the embedded clues.

The following examples show items and how they were altered to include clues that could elicit test-wise behaviors. In each case, the test-wiseness skills were derived from Sarnacki (1979, p. 255).

1. The student demonstrates the ability to eliminate options which are known to be incorrect and choose from the remaining options. An example of demonstrating this behavior would be to exclude option "a" in the following item:

Which of these animals is a mammal?
a. a rock
b. a dog
c. a snake
d. a bird

Option "a" can be eliminated from consideration because it is absurd.

2. The student demonstrates the ability to choose neither one of two options that overlap or are synonymous to each other. An example of this behavior would be to exclude options "b" and "d" from consideration in the following item:

Which of these animals can fly?
a. a bat
b. a snake
c. a rabbit
d. a cobra

Options "b" and "d" overlap each other since a cobra is a kind of snake. The student demonstrating this test-wise skill would then choose from options "a" and "c."

3. The student demonstrates the ability to recognize that often the correct option is the one that is longer or has more detail than other options. In this case, the student demonstrating this behavior would tend to select option "c" in the following item:

Which of these statements best describes salt?
a. It is a food preservative.
b. It tastes good.
c. It is a compound consisting of a sodium atom and a chlorine atom.
d. It helps melt snow in the winter.

Option "c" is twice as long as any other option and includes more specific detail. Of course, test developers also use this type of option as a distractor so it is not always the correct option. This item without clues might be:

Which of these statements best describes salt?
a. It is a food preservative.
b. It tastes good.
c. It is a compound of sodium and chlorine.
d. It helps melt snow in the winter.

4. The student demonstrates the ability to recognize the lack of grammatical agreement between the stem and incorrect options or recognizes a grammatical hint in the correct option. The following item serves as an example, and the test-wise student would tend to select option "d."

The largest mammal is a _____
a. elephant
b. antelope
c. elk
d. whale

Notice that the stem ends with "a" instead of "an" which suggests to the test-wise student that the first three options can be excluded because they begin with a vowel sound.

5. The student demonstrates the ability to recognize a verbal association between the stem and one of the options and then tends to select this option as the correct option. The following item includes a verbal association or hint to the correct response. This item is also an example of a pseudo-item since it contains imaginary or fictitious content. In this item, option "a" is closely aligned with the content of the stem and would tend to be chosen by a test-wise student.

In which city is the Newmunichopola located?
a. Munich
b. New York
c. Paris
d. San Francisco

The test-wise student would select option "a" because Munich is the only city on the list that is verbally associated with the word "Newmunichopola."

Design

A correlational design was used in this study. Subscores on each of the five test-wiseness skills were computed for each examinee. These five subscores were computed for both the clued and the unclued forms of the test.

Procedures

The testing was completed during a two-day period with one day being spent in each school. Tests were administered to each grade-level independently. Detailed instructions were read to students in each grade, and researchers and teachers then monitored the testing session.

A short break was given between each group of 30 items so that students started working on each group of 30 items at the same time. The purpose of the break was to help avoid fatigue and to help keep students from working on other sections of the test. Students

who finished a section early were instructed to read a library book while waiting for other students to finish the section. The test was not timed since it was desirable that each student be able to complete each question. Since the clued items were all in one section of the test, it was necessary to make sure students did not go back and change answers in any previous section. Students were instructed not to go back and change answers once they had started with the next set of items.

To control for the possibility of an order or sequence effect, half of the students in each grade completed the unclued section followed by the clued section of the test. The other half of the students completed the clued section first. Students were randomly assigned to this testing sequence.

Analysis

All items were scored dichotomously. Since items were originally written with right-wrong scoring in mind, partial credit scoring was not used (Millman and Greene, 1989, pp. 343-344). Subscale scores for each of the specific types of test-wiseness were computed by counting the number of items answered correctly for each skill. Subscale scores were also computed for the pseudo-items. Reliability estimates were calculated for each of the subscales. For subscales originally composed of more than five items, items with the lowest correlations with other items in the subscale were removed. This procedure maximized the reliability estimates and provided an equal number of items in each subscale. Ultimately, each subscale consisted of five items.

Originally, a repeated measures ANOVA was used to compare the students' mean performance on the clued and unclued version of each subscale respectively. Testing sequence and sex were included in the model as independent variables, but in each analysis, no statistically significant main effect for testing sequence was found, ($p > .05$).

In each case where a statistically significant main effect for sex was found, the effect was the same for both clued subscales and unclued subscales. Consequently, the data were re analyzed using paired sample t-tests as a simpler statistical model.

For fifth- and sixth-grade students, the subscale score distributions on the pseudo-items were then used to classify students as either masters or non-masters in terms of demonstrating the test-wiseness skills. A cut-off of 60% correct was used on all five scales. This figure corresponded to students correctly answering at least three of the five items in each subscale. This mastery/non-mastery classification was used to help compare the developmental patterns of test-wiseness skills of fifth- and sixth-grade students. Independent sample t-tests were used for each comparison between fifth and sixth-grade students as well as for comparisons by sex.

Results

The analyses were completed within grade-level except for comparisons made between fifth- and sixth-grade students on the pseudo-item subscales.

Third Grade

Table 1 reports the means, standard deviations, and reliability estimates of the clued and unclued subscales designed to assess the various test-wiseness skills among third graders. Each of the reported reliability coefficients is quite low and suggests a lack or consistency in the students' responses to the individual items that make up each subscale. For the clued subscales the alpha coefficients ranged from .24 to .64, and for the unclued subscales, the coefficients ranged from .04 to .63.

Insert Table 1 about here

As Anastasi suggests, coefficient alpha is often used as a means of estimating the reliability of scores obtained from a single administration of a single test, but it is more precise to think of this coefficient as an index of "interitem consistency" or homogeneity (1988, p. 122). She points out that the size of this index is influenced by two sources of error variance: (a) the heterogeneity of the subject-matter content of the items sampled; and (b) the heterogeneity of the behavior domain sampled. The items in each subscale were not selected on the basis of similar content or mental tasks, but for ease in embedding clues. Since the items were not homogeneous in content, one should not expect to obtain a high index of inter-item consistency unless students were consistently attending to the embedded clues in the clued items. If students were attending to the clues, the alpha-coefficients for the clued subscales would be higher and would be larger than the coefficients for the unclued subscales.

The third-grade students did not correctly answer more clued items than unclued items. This finding applied to both boys and girls in the third grade. The five paired t-tests indicated a statistically significant difference in the means of the clued and unclued versions on only one of the test-wiseness skills, the subscale which assessed students' ability to recognize synonymous or overlapping options and to use this knowledge as a basis for identifying the correct answer by ruling out the synonymous choices. In this case, the third graders correctly answered more unclued items than clued items.

Table 2 summarizes the intercorrelation coefficients between each of the five clued subscales and each of the five unclued subscales. Note that the highest coefficients in the matrix occur in the diagonal shown in boldface type and range from .68 to .83. The coefficients in this diagonal represent the correlations between the clued and unclued versions of the five subscales. Hence, each correlation in this diagonal is between

subscales consisting of items that differ from each other only in the presence or absence of embedded clues. The science content of the items in each of the respective subscales was identical.

Insert Table 2 about here

On the other hand, the correlation coefficients between the other clued and unclued subscales are only moderate in size and range from .22 to .52. For the sample size reported here, each of these correlation coefficients is significantly different from zero, ($p < .01$).

However, for each subscale, the coefficients in the diagonal are significantly higher than those in the off-diagonal, ($p < .05$). Items in the subscales along the diagonal differed only in the presence or absence of clues. The science content was the same. In the off-diagonal, the items differed not only in the presence or absence of clues, but also in the science content of the items suggesting that it was the science content of the items rather than attention to the clues in the items that was the key factor in the responses of third graders. When the science content was not similar, the correlation coefficients were attenuated. Of course, another possible explanation for the attenuation of the correlation coefficients is the restriction of the range due to the poor performance of the students due to the non-functioning of the clues.

The intercorrelations between the clued subscales are moderate in size and range from .38 to .54. The intercorrelations between the unclued subscales are also moderate and range from .31 to .47. Although the correlation coefficients are moderate, this result

was expected since the subscales are intended to assess different skills. The correlation coefficients are significant ($p < .01$) for the sample size reported here.

As a group then, the third-grade students did not demonstrate proficiency in any of the test-wise behaviors that were measured. They did not make use of the clues embedded in the items and apparently relied on some combination of their knowledge of the content and random response as they answered each question.

Fifth Grade

As a group, the fifth-grade students answered more items correctly when they had clues than when the items were unclued. The five paired t-tests revealed statistically significant differences between the mean scores on four of the five scale pairs, all except the the grammatical inconsistencies pair. This finding applied to both boys and girls and suggests that fifth-grade students tended to notice the clues and that the presence of the clues tended to help them select the correct answer.

Table 3 reports the descriptive statistics and reliability estimates for the clued and unclued versions of the respective subscales for fifth-grade students. The estimated reliability coefficients are quite low and suggest little consistency in the responses of students to the individual items on each scale. The alpha coefficients for the clued subscales ranged from only .09 to .45, and for the unclued subscales from .23 to .32 suggesting (a) that the items were not homogeneous in content and (b) that students did not consistently attend to the clues in the clued items.

Insert Table 3 about here

Table 4 summarizes the intercorrelations between each of the clued scales and each of the unclued scales. The size of the correlation coefficients between scale scores are low as one would expect for scales intended to assess different skills. The highest correlations are between the respective pairs of scales consisting of clued and unclued versions of the same items (the correlation coefficients shown in bold-faced type in Table 4) but even these correlation coefficients are moderate in size. The correlations between different skills are even lower for fifth-grade students and the coefficients range from .11 to .38 suggesting that the relationship between the different test-wiseness skills tend to be rather weak. For the sample size reported here, correlation coefficients of .20 and greater are significantly different from zero, ($p < .01$).

Insert Table 4 about here

However, unlike the correlation coefficients among the third graders, the coefficients on the diagonal are not all significantly different from those on the off-diagonal. The correlation between the clued scale assessing the verbal association skill and its companion unclued scale was not significantly greater than the correlation between this clued scale and other unclued scales, ($p > .05$). This same result was found for the absurd options subscale. These findings indicate that the science content of the items alone does not account for students' responses. If the students had been attending only to the science content of the items, one would expect the correlation coefficients to be somewhat higher. On the other hand, if students attend to the clues and thereby obtain higher scores, one would expect a weaker relationship between the clued and unclued subscales.

Consequently, it appears that some fifth graders are attending to the clues in the clued items thereby attenuating the correlation coefficients among similar items.

As expected, since they are intended to assess different skills, the intercorrelations between the clued subscales are only moderate in size and range from .16 to .35, and the intercorrelations between the unclued subscales range from .11 to .38. Correlation coefficients of .20 and greater are significant ($p < .01$) given the sample size reported here.

Overall, some fifth-grade students demonstrated test-wise behaviors on individual test-wiseness skills. Some fifth-grade students attended to clues related to (a) verbal associations, (b) absurd options, (c) synonymous or overlapping options, or (d) longer or more detailed options. However, the skills are not correlated with each other. This finding suggests that students who demonstrate one test-wise behavior are not likely to demonstrate other test-wise behaviors. In addition, fifth-grade students have developed test-wise behaviors not evident among third-grade students.

Sixth Grade

As a group, the sixth-grade students correctly answered more clued items than unclued items. The five paired t-tests revealed statistically significant differences between the mean scores on three of the five scales. The two exceptions were the scale assessing the grammatical inconsistencies skill and the scale assessing the synonymous or overlapping options skill. This finding applied to both boys and girls and suggests that sixth-grade students demonstrated some test-wise behaviors. In other words, the test-wiseness skills developed by the fifth grade do not change dramatically in the sixth grade and appear to be continuing from the fifth grade.

Table 5 reports the means, standard deviations, and reliability estimates for the five clued and the five unclued scales for sixth-grade students. Just like the alpha coefficients for the scales among third- and fifth-grade students, the reliability coefficients for the sixth-grade students are low and indicate inconsistency in the responses of students to the individual items in each scale. The alpha coefficients for the clued subscales ranged from .28 to .53, and the coefficients for the unclued subscales ranged from .32 to .60 suggesting that many sixth graders, like fifth graders, did not consistently attend to the embedded clues. This finding suggests that the clues may function in a way that is item-specific. In other words, students may identify the clue in one item but overlook a similar clue in another item.

Insert Table 5 about here

Table 6 summarizes the intercorrelations between the each of the five clued scales and each of the five unclued scales among sixth graders. These correlations between scales are modest, but somewhat higher than those reported for the fifth grade. As expected, the highest correlations are found between the clued and unclued versions of the scales containing items that differed only in the presence or absence of embedded clues (the correlation coefficients appearing in bold-faced type in Table 6). These correlations were relatively high for sixth-grade students. On the other hand, the correlations between different test-wiseness skills and between the different clued and unclued scales range from .27 to .59 indicating a moderate relationship between the different test-wiseness skills. Correlation coefficients of .20 and greater are significantly different from zero, ($p < .01$), for the sample size reported here.

Insert Table 6 about here

On the other hand, like the fifth graders, the sixth graders' responses appeared to be based not only on the science content of items but also on the embedded clues. The correlation between the clued verbal association subscale and its companion scale of unclued items was not significantly greater than the correlations between the clued verbal association subscale and other unclued subscales ($p > .05$).

Comparisons between Fifth and Sixth-Grade Students

Table 7 displays the means, standard deviations, and reliability estimates for the five pseudo-item subscales for fifth and sixth graders. The reliability coefficients are relatively low indicating a lack of consistency in the responses of both fifth- and sixth-grade students to the individual items on each scale. The alphas for the synonymous or overlapping options subscales among both fifth and sixth graders were especially low. A negative alpha (-.06) was calculated for fifth graders on this scale, an indication of negative covariance between some of the items. The findings related to the synonymous or overlapping options subscale suggest that students in the fifth and sixth grade did not attend to the clues in a consistent manner.

Insert Table 7 about here

However, for each subscale, the alpha coefficient for the sixth graders was higher than the alpha coefficient for fifth graders. This finding suggests that sixth-grade students were more consistent in responding to the individual items within each scale.

Comparison of the means of fifth graders with the means of sixth graders on the corresponding subscales provides further evidence that sixth-grade students demonstrated more test-wiseness than fifth-grade students. Sixth graders made better use of the clues embedded in the pseudo-items than fifth-graders on (a) the verbal associations subscale and (b) the absurd options subscale suggesting that they not only tend to maintain the test-wiseness skills learned by the fifth grade, but also tend to strengthen their use in the sixth grade. Sixth graders also performed significantly better than fifth graders on the synonymous or overlapping options subscale, but given the low alpha coefficients this finding may be a statistical artifact.

Table 8 shows the intercorrelations between the five pseudo-item subscale scores for both the fifth graders and the sixth graders. Each of these correlation coefficients is low, suggesting that the subscales assess different and independent traits. However, among sixth-grade students, there is a moderate relationship (.45) between the verbal associations subscale and the absurd options subscale. This relationship suggests that sixth-grade students who identify absurd options also have a tendency to identify verbal associations and attend to both types of clues. However, this relationship was not found for fifth-grade students, suggesting that test-wiseness skills develop independently of each other.

Insert Table 8 about here

The students' scores on each of the pseudo-item subscales were used as a basis for classifying them as having mastered or not mastered each of the skills respectively. Table 9 summarizes the percent of students in each grade who were classified as masters on each

subscale. The pseudo-item subscale scores provide evidence of the development of two of the test-wiseness skills among fifth-grade students. Some fifth-grade students made use of verbal association clues or absurd option clues when responding to multiple-choice test items.

Insert Table 9 about here

The pseudo-item subscale scores also provide evidence of the development of three of the test-wiseness skills among sixth-grade students. Sixth-grade students demonstrated a well-developed ability to use the absurd options clues and the verbal association clues in responding to pseudo-items. Some sixth-grade students also demonstrated the development of skill in using synonymous or overlapping options clues.

In general, both fifth- and sixth-grade students made some use of clues embedded in pseudo-items, but among sixth graders, test-wiseness appears to have been more widespread. Table 10 shows that many more sixth-grade students than fifth-grade students had mastered one or more test-wiseness skills. Approximately 14% of the fifth graders had not mastered any of the five skills, and a third of them demonstrated mastery on only one skill. On the other hand, only 3% of the sixth graders had not mastered any of the five skills, and 18% had mastered only one of the five skills. Few students in either grade had mastered more than three of the five test-wiseness skills. Not only does test-wiseness appear to be developmental in nature, but the various test-wiseness skills tend to develop independently of each other.

Insert Table 10 about here

A sex comparison for both fifth- and sixth-grade students indicated that sex is not an important factor in the development of test-wiseness skills. Table 11 summarizes the descriptive statistics and the results of the t-test comparisons by sex for fifth graders. Table 12 summarizes the descriptive statistics and gives the results of the t-test comparisons by sex for the sixth graders. No significant differences were found between boys and girls for any of the five pseudo-item scales for either fifth or sixth graders.

Insert Table 11 about here

Insert Table 12 about here

Discussion

The purpose of this study was to investigate the prevalence of test-wiseness skills among third-, fifth-, and sixth-grade students. Comparisons between grades were limited to those between fifth- and sixth-grade students on similar pseudo-items. Since the science tests were grade-level specific and because of differences in the reading abilities and maturity of third-, fifth-, and sixth-grade students, other comparisons among the grades were not possible.

Third-grade students showed little development of any of the five test-wiseness skills that were investigated. No differences were found between their performance on the

clued items and unclued items that would indicate test-wise behavior. Reliability estimates suggested that third-grade students did not attend to the embedded clues and were inconsistent when responding to both the clued and the unclued items. Neither boys nor girls attended to the clues for any of the test-wiseness skills.

On the other hand, many of the fifth-grade students demonstrated the ability to take advantage of embedded clues to increase their test scores. Both boys and girls performed better on four of the five clued subscales than they did on the unclued subscales. However, the wide ranges observed in the reliability estimates suggests that fifth-grade students were also inconsistent in their responses to both clued and unclued items when they were uncertain of the best response.

Fifth-grade students also demonstrated a sensitivity to clues when responding to pseudo-items. Responses of students indicated development of two of the five test-wiseness skills that were measured using pseudo-items. However, the range of the alpha coefficients and the moderate size of these coefficients suggests once again that fifth graders responded inconsistently.

In addition, fifth-grade students appeared to be more likely to benefit from clues when actual items rather than pseudo-items were used. They averaged higher scores on four of the five clued subscales, but only demonstrated mastery on two of the five pseudo-item subscales. Apparently, the science content and the student's knowledge of that content had some effect on the attention fifth graders gave to the embedded clues. Since previous research has used mainly pseudo-items or items unrelated to the content of the curriculum to collect information on test-wise behaviors, more research is needed to investigate the effects of clues embedded in items directly related to the content being tested.

Both boys and girls in the sixth-grade demonstrated the ability to identify and successfully use clues to increase their scores on three of the five clued subscales and on three of the five pseudo-item subscales. On each of the three pseudo-item subscales in which they demonstrated test-wise behaviors, sixth-grade students outperformed the fifth-grade students. Sixth graders continued to develop test-wiseness skills and were able to make more use of the clues than fifth graders. On the other hand, not all students have developed test-wiseness skills by the sixth grade. The moderate reliability estimates and the range of the alpha coefficients on the clued subscales and the pseudo-item subscales suggest that sixth-grade students were also inconsistent in attending to embedded clues. In addition, they responded inconsistently to the unclued items.

However, the alpha coefficients were higher on each of the pseudo-item subscales for sixth graders than for fifth graders suggesting more consistency in responses and more identification and use of embedded clues by sixth-grade students. This finding is further evidence of the developmental nature of test wise behaviors.

In addition, the inconsistent responses of students to items with similar clues suggests that the effect of clues is item specific or context dependent. In other words, a student may identify and make use of the clue in one item but fail to recognize a similar clue in another item. This finding was observed not only in the responses of fifth- and sixth-grade students to pseudo-items, but also in their responses to items in the clued section of the test. Diamond and Evans (1972) also found that test-wiseness skills tend to be content dependent. An important follow-up to this study would be to investigate the cognitive processes of students as they respond to both pseudo-items and other clued items to try to determine: (a) why clues related to a specific test-wiseness skill operate so

differently among individual students; and (b) how the function of clues differs from one context to another.

The grammatical inconsistencies skill was the least developed skill among the students in the third, fifth, or sixth grade. Few students attended to the grammatical clues in either the clued items or the pseudo-items. A plausible explanation is that the students had not yet been taught the appropriate grammatical rules. Most of the clues were based on (a) the correct use of "a" or "an" or (b) subject-verb agreement, and such topics are often not introduced to the students until later grades.

Additional research is also needed to investigate why some skills develop earlier and to examine how these skills develop cognitively. As part of this new research, effective ways of comparing skill development among students in earlier grades (first, second, and third) with those in the later elementary grades (fourth, fifth, and sixth) are needed. Differences in reading level, cognitive development of the students, and content domain of the tests are issues that must all be addressed when making comparisons across grades.

Sex was not related to the development of test-wiseness skills. Both boys and girls in the fifth grade and the sixth grade demonstrated the ability to attend successfully to clues embedded in items. However, additional research is needed to investigate the relationship between test-wiseness skills in elementary school students and other demographic variables. Furthermore, the intercorrelations between the various subscales indicated little relationship between the five test-wiseness skills that were measured in this study and suggest the five test-wiseness skills develop independently of each other. This lack of inter-relationship was observed in all three of the grades tested. Test-wiseness is not a single construct, but consists of a number of independent skills. However, there is a slight indication that two of the test-wiseness skills, skill in

identifying verbal associations and skill in recognizing absurd options are moderately correlated among sixth graders. Therefore, more research is needed to further investigate the independent development of skills to determine the extent to which test-wiseness skills become more inter-related as students develop cognitively.

How students acquire test-wiseness skills was not a focus of this study, but it is apparent that as students mature and move from grade to grade more skills are developed. The first skills to be developed in the fifth grade are usually those related to identifying a verbal association between the stem and one of the response options or recognizing response options that are absurd. These two skills tend to be followed in the sixth grade by the skill of recognizing that two or more response options overlap or are synonymous to each other. Recognizing grammatical inconsistencies and being attracted to longer or more detailed response options are test-wisenass skills not developed to a great extent among elementary school students. Apparently, test-wiseness skills begin to appear as children begin to move into adolescence and develop more formal logical structures. Additional research is needed to investigate how test-wiseness skills develop among elementary school students and why some students develop certain skills before other students. General scholastic attitude of students would be an important factor to include in such future studies.

Finally, most previous test-wiseness research has depended on the use of pseudo-items or nonsense items to measure levels of test-wiseness. However, more research is needed using items from existing tests paired with altered forms of each item in addition to pseudo-items. Item analysis procedures could be used to investigate the differences in the ways the various response options are selected for clued and unclued items. As part of this item analysis, the process of distractor analysis proposed by Millman, and Greene (1989,

pp. 361-362) could help identify the extent to which distractors function differently when they are clued or unclued. In addition, the distractors may function differently for different groups of students. For example, if some word or phrase in an item inadvertently provides a clue to one group of students that helps them obtain higher test scores than another group of students, evidence would be found to suggest the item was biased. Consequently, in future research on item bias, test-wiseness skills should be included as part of the study. Differences in test-wiseness may help explain between-group differences usually interpreted as item bias.

References

Anastasi, A. (1988). Psychological testing, (6th ed.). New York: Macmillan.

Brainerd, C. J. (1978). Piaget's theory of intelligence. Englewood Cliffs, NJ: Prentice-Hall.

Case, R. (1985). Intellectual development: Birth to adulthood. Orlando, FL: Academic Press, Inc.

Diamond, J. J., & Evans, W. J. (1972). An investigation of the cognitive correlates of test-wiseness. Journal of Educational Measurement, 9, 145-150.

Ebel, R.L. & Frisbie, D.A. (1986). Essentials of educational measurement, (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Gronlund, N.E. (1985). Measurement and evaluation in teaching (5th ed.). New York: Macmillan.

Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. Applied Measurement in Education, 2, 37-49.

McMorris, R. F., Brown, J. A., Snyder, G. W., & Pruzek, R. M. (1972). Effects of violating item construction principles. Journal of Educational Measurement, 9, 287-295.

Mehrens, W.A. & Lehmann, I.J. (1984). Measurement and evaluation in education and psychology (3rd ed.). New York: Holt, Rinehart & Winston.

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-726.

Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), Educational Measurement. Washington, DC: American Council on Education.

Parrish, B. W. (1982). A test to test test-wiseness. Journal of Reading, 26, 672- 675.

Popham, W.J. (1981). Modern educational measurement. Englewood Cliffs, NJ: Prentice-Hall.

Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. Review of Educational Research, 49, 252-279.

Siegler, R. S. (1986). Children's thinking. Englewood Cliffs, NJ: Prentice-Hall.

Slakter, M. J., Koehler, R. A., & Hampton, S. H. (1970). Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 7, 119-122.

Scruggs, T. E., Bennion, K., & Lifson, S. (1985). An analysis of children's strategy use on reading achievement tests. Elementary School Journal, 85, 479-484.

Scruggs, T. E., White, K. R., & Bennion, K. (1986). Teaching test-taking skills to elementary-grade students: A meta-analysis. Elementary School Journal, 87, 69-82.

Utah State Board of Education (1987). Elementary and secondary core curriculum. Salt Lake City: Utah State Office of Education.

Table 1

Descriptive Statistics and Reliability Estimates of Subscales for Third-Grade Students

Scale Description	N	Number of Items	Alpha Coefficient	Mean Score	SD	t Value	df	Significance Level
Grammatical Inconsistencies								
Clued Items	134	5	.30	2.95	1.25			
Unclued Items	134	5	.38	2.96	1.25	-.19	133	.848
Verbal Associations								
Clued Items	134	5	.32	3.46	.97			
Unclued Items	134	5	.04	3.50	.86	-.58	133	.560
Synonymous or Overlapping Options								
Clued Items	134	5	.37	3.84	1.07			
Unclued Items	134	5	.32	3.94	1.01	-1.47	133	.145
Absurd Options								
Clued Items	134	5	.64	4.10	1.20			
Unclued Items	134	5	.63	4.25	1.13	-2.44	133	.016
Longer or More Detailed Options								
Clued Items	134	5	.24	3.33	1.12			
Unclued Items	134	5	.35	3.29	1.18	.60	133	.549

Table 2

Intercorrelations between Clued and Unclued Subscale Scores among Third-Grade Students

	Clued Subscales					Unclued Subscales				
	C1	C2	C3	C4	C5	U1	U2	U3	U4	U5
Clued Subscales										
(C1) Grammatical Inconsistencies	--	.42	.38	.44	.37	.74	.38	.31	.43	.34
(C2) Verbal Associations		--	.42	.54	.41	.39	.68	.39	.47	.34
(C3) Synonymous Options			--	.46	.36	.39	.34	.69	.41	.41
(C4) Absurd Options				--	.43	.51	.52	.42	.83	.44
(C5) Longer Options					--	.30	.35	.22	.43	.80
Unclued Subscales										
(U1) Grammatical Inconsistencies						--	.37	.31	.47	.27
(U2) Verbal Associations							--	.35	.45	.33
(U3) Synonymous Options								--	.41	.31
(U4) Absurd Options									--	.38
(U5) Longer Options										--

Note. Column headings are the same as row headings.

Table 3

Descriptive Statistics and Reliability Estimates of Subscales for Fifth-Grade Students

Scale Description	N	Number of Items	Alpha Coefficient	Mean Score	SD	t Value	df	Significance Level
Grammatical Inconsistencies								
Clued Items	133	5	.20	3.15	1.15			
Unclued Items	133	5	.27	3.11	1.19	.40	132	.691
Verbal Associations								
Clued Items	133	5	.09	2.02	1.09			
Unclued Items	133	5	.29	1.74	1.15	2.66	132	.009
Synonymous or Overlapping Options								
Clued Items	133	5	.38	2.65	1.27			
Unclued Items	133	5	.32	2.38	1.21	2.73	132	.007
Absurd Options								
Clued Items	133	5	.45	3.07	1.30			
Unclued Items	133	5	.26	2.53	1.16	5.26	132	.000
Longer or More Detailed Options								
Clued Items	133	5	.42	2.55	1.27			
Unclued Items	133	5	.23	2.20	1.11	3.62	132	.001

Table 4

Intercorrelations between Clued and Unclued Subscale Scores among Fifth-Grade Students

	Clued Subscales					Unclued Subscales				
	C1	C2	C3	C4	C5	U1	U2	U3	U4	U5
Clued Subscales										
(C1) Grammatical Inconsistencies	--	.34	.35	.23	.16	.57	.26	.29	.21	.15
(C2) Verbal Associations		--	.28	.24	.33	.18	.42	.32	.20	.24
(C3) Synonymous Options			--	.31	.24	.34	.19	.55	.26	.21
(C4) Absurd Options				--	.31	.26	.15	.43	.52	.28
(C5) Longer Options					--	.16	.25	.17	.21	.58
Unclued Subscales										
(U1) Grammatical Inconsistencies						--	.20	.37	.35	.26
(U2) Verbal Associations							--	.17	.25	.11
(U3) Synonymous Options								--	.38	.25
(U4) Absurd Options									--	.26
(U5) Longer Options										--

Note. Column headings are the same as the row headings.

Table 5

Descriptive Statistics and Reliability Estimates of Subscales for Sixth-Grade Students

Scale Description	N	Number of Items	Alpha Coefficient	Mean Score	SD	t Value	df	Significance Level
Grammatical Inconsistencies								
Clued Items	143	5	.28	2.90	1.22			
Unclued Items	143	5	.32	2.93	1.25	-.35	142	.729
Verbal Associations								
Clued Items	143	5	.40	2.86	1.33			
Unclued Items	143	5	.37	2.55	1.30	2.99	142	.003
Synonymous or Overlapping Options								
Clued Items	143	5	.53	2.77	1.44			
Unclued Items	143	5	.60	2.69	1.51	.99	142	.323
Absurd Options								
Clued Items	143	5	.33	3.52	1.13			
Unclued Items	143	5	.43	3.19	1.23	3.96	142	.000
Longer or More Detailed Options								
Clued Items	143	5	.50	3.41	1.29			
Unclued Items	143	5	.36	3.21	1.20	2.50	142	.014

Table 6

Intercorrelations between Clued and Unclued Subscale Scores among Sixth-Grade Students

	Clued Subscales					Unclued Subscales				
	C1	C2	C3	C4	C5	U1	U2	U3	U4	U5
Clued Subscales										
(C1) Grammatical Inconsistencies	--	.35	.55	.53	.31	.70	.42	.55	.50	.37
(C2) Verbal Associations		--	.44	.27	.38	.27	.54	.49	.31	.29
(C3) Synonymous Options			--	.59	.53	.48	.38	.80	.48	.46
(C4) Absurd Options				--	.35	.44	.29	.50	.63	.35
(C5) Longer Options					--	.34	.35	.50	.36	.72
Unclued Subscales										
(U1) Grammatical Inconsistencies						--	.43	.48	.51	.41
(U2) Verbal Associations							--	.48	.40	.42
(U3) Synonymous Options								--	.53	.48
(U4) Absurd Options									--	.42
(U5) Longer Options										--

Note. Column headings are the same as row headings.

Table 7

Descriptive Statistics and Reliability Estimates of the Pseudo-Item Subscale Scores for Fifth- and Sixth-Grade Students

Scale Description	N	Number of Items	Alpha Coefficient	Mean Score	SD	t Value	df	Significance Level
Grammatical Inconsistencies								
Fifth Graders	133	5	.42	1.49	1.24			
Sixth Graders	143	5	.56	1.79	1.38	1.91	274	.057
Verbal Associations								
Fifth Graders	133	5	.29	2.44	1.19			
Sixth Graders	143	5	.43	2.87	1.28	2.88	274	.004
Synonymous or Overlapping Options								
Fifth Graders	133	5	-.06	2.08	1.06			
Sixth Graders	143	5	.13	2.60	1.17	3.90	274	.000
Absurd Options								
Fifth Graders	133	5	.30	2.63	1.27			
Sixth Graders	143	5	.52	3.78	1.23	7.64	274	.000
Longer or More Detailed Options								
Fifth Graders	133	5	.29	1.77	1.19			
Sixth Graders	143	5	.42	1.69	1.25	.61	274	.545

Table 8

Intercorrelations between Pseudo-item Subscale Scores for Pseudo-Items for Fifth- and Sixth-Grade Students

	Pseudo-Item Subscales				
Subscales	P1	P2	P3	P4	P5
Fifth Graders					
(P1) Grammatical Inconsistencies	--	.11	.19	.14	.05
(P2) Verbal Associations		--	.10	.13	.11
(P3) Synonymous Options			--	.08	-.14
(P4) Absurd Options				--	.01
(P5) Longer Options					--
Sixth Graders					
(P1) Grammatical Inconsistencies	--	.27	.17	.34	.02
(P2) Verbal Associations		--	.28	.45	.11
(P3) Synonymous Options			--	.20	.11
(P4) Absurd Options				--	-.14
(P5) Longer Options					--

Note. Column headings are the same as row headings.

Table 9

Number and Percent of Fifth and Sixth-Grade Students Demonstrating Mastery by Type of Test-Wiseness Skill

	Fifth Graders		Sixth Graders	
Test-Wiseness Skill	N	Percent	N	Percent
Absurd Options	65	49%	122	85%
Verbal Associations	60	45%	89	62%
Synonymous or Overlapping Options	49	37%	74	52%
Longer or More Detailed Options	33	25%	34	24%
Grammatical Inconsistencies	20	15%	34	24%

Table 10

Number of Test-Wiseness Skills Mastered by Grade-Level

	Fifth Grade		Sixth Grade	
Number of Skills Mastered	N	Percent	N	Percent
None	19	14%	4	3%
One	44	33%	26	18%
Two	39	29%	49	34%
Three	19	14%	32	22%
Four	12	9%	27	19%
Five	0	0	5	4%
Totals	133	100%	143	100%

Table 11

Descriptive Statistics for Pseudo-Item Subscales by Sex among Fifth-Grade Students

Scale Description	N	Mean Score	SD	t Value	d.f.	Significance Level
Grammatical Inconsistencies						
Boys	67	1.66	1.40			
Girls	65	1.34	1.02	1.49	130	.139
Verbal Associations						
Boys	67	2.34	1.21			
Girls	65	2.55	1.17	-1.01	130	.313
Synonymous or Overlapping Options						
Boys	67	2.07	1.18			
Girls	65	2.06	.93	.07	130	.944
Absurd Options						
Boys	67	2.58	1.37			
Girls	65	2.68	1.17	-.43	130	.671
Longer or More Detailed Options						
Boys	67	1.81	1.06			
Girls	65	1.74	1.33	.32	130	.747

Table 12

Descriptive Statistics for Pseudo-Item Subscales by Sex among Sixth-Grade Students

Scale Description	N	Mean Score	SD	t Value	d.f.	Significance Level
Grammatical Inconsistencies						
Boys	88	1.91	1.41			
Girls	55	1.60	1.31	1.31	141	.193
Verbal Associations						
Boys	88	2.82	1.33			
Girls	55	2.96	1.22	-.66	141	.511
Synonymous or Overlapping Options						
Boys	88	2.60	1.27			
Girls	55	2.60	.99	.01	141	.991
Absurd Options						
Boys	88	3.86	1.22			
Girls	55	3.65	1.27	.99	141	.326
Longer or More Detailed Options						
Boys	88	1.66	1.26			
Girls	55	1.73	1.24	-.32	141	.752